

[First Hit](#)[Previous Doc](#)[Next Doc](#)[Go to Doc#](#) [Generate Collection](#) [Print](#)

L11: Entry 12 of 27

File: JPAB

Nov 30, 1999

PUB-NO: JP411328275A

DOCUMENT-IDENTIFIER: JP 11328275 A

TITLE: METHOD AND SYSTEM FOR INFORMATION FILTERING

PUBN-DATE: November 30, 1999

INVENTOR-INFORMATION:

NAME

COUNTRY

NAKAMURA, ATSUYOSHI

ASSIGNEE-INFORMATION:

NAME

COUNTRY

NEC CORP

APPL-NO: JP10150593

APPL-DATE: May 14, 1998

INT-CL (IPC): G06 F 17/60; G06 F 17/18

ABSTRACT:

PROBLEM TO BE SOLVED: To provide information filtering method and system which consider reliability because of making inter-user weight correlation coefficient, enable inter-user weighting so as to make the weight difference between inter-users who are similar and inter-users who are not larger and improve estimation performance due to that.

SOLUTION: This system has an estimating means 3 which defines a set of values that votes for each candidate value and decides an estimation value by majority due to the total of weights among users of evaluated values that enters the set and a learning means 4 which calculates the weights among the users by calculating the correct answer degree and error degree of estimation among the users from a number voting for a correct answer and a number voting for an incorrect answer, multiplying 1 by a number that is larger than 1 as many as the value of the correct answer degree and multiplying it by a number that is smaller than 1 as many as the number of the error degree and performs weight learning among the users by storing calculated values or storing the number of correct answers and errors.

COPYRIGHT: (C)1999, JPO

[Previous Doc](#)[Next Doc](#)[Go to Doc#](#)

7

重みの合計sumとする(ステップ34201)。

【0024】図5は、本発明の第1の実施例における学習手段4の処理フローを示す流れ図である。正解値M(i, j)に投票した行i' (O(i', j) ∈ V(M(i, j)))と、行iとの間の重みw(i, i')は、(2-γ)倍にし(ステップ450)、M(i, j)に投票しない他の値に投票した行i' (O(i', j))が「?」でなく、しかもV(M(i, j))に含まれていないと行iとの間の重みw(i, i')はγ倍にする(ステップ460)。但し、γは0以上1未満の定数である。

【0025】重みの更新後、観測行列の(i, j)エントリの値O(i, j)を、M(i, j)に更新する(ステップ48)。

【0026】[実施例2]図6乃至図8は、本発明の第2の実施例の処理フローを説明するための流れ図である。図6乃至図8を参照して本発明の第2の実施例の動作について詳細に説明する。

【0027】図6は、本発明の第2の実施例における初期化手段2の処理フローを示す流れ図である。任意のユーザiとユーザi'間の正しい投票回数C(i, i')と間違った投票回数D(i, i')を「0」にセットし(ステップ231)、その時までに観測した行列エントリの値((i, j)成分はユーザiによる情報jの評価値)を記憶するために使う観測行列の任意のエントリ(i, j)の値O(i, j)を「?」にセットする(ステップ24)。

【0028】本発明の第2の実施例における予測手段3の処理は、図3に示した第1の実施例における予測手段3の処理フローと同じである。

【0029】図7は、本発明の第2の実施例において、図3のsumの計算をするステップ34の処理フローを示す流れ図である。候補値aに投票する値の集合をV(a)とする。同じ列jで値が既に分かっている行i'の値O(i', j)がV(a)に含まれているような行i' と行iとの間の重みpow(2-γ, C(i, i'))pow(γ, D(i, i'))の合計を、候補値aに投票する重みの合計sumとする(ステップ34211)。但しpow(x, y)はxのy乗を表す。

【0030】図8は、本発明の第2の実施例における学習手段4の処理フローを示す流れ図である。正解値M(i, j)に投票した行i' (O(i', j) ∈ V(M(i, j)))と行iとの間の正しい投票回数を数えるカウンタC(i, i')に1を加え(ステップ451)、M(i, j)に投票しない他の値に投票した行i' (O(i', j))は「?」でなくしかもV(M(i, j))に含まれていないと行iとの間の間違った投票回数を数えるカウンタD(i, i')に「1」を加える(ステップ461)。重みの更新後、観測行列の(i, j)エントリの値O(i, j)をM(i, j)に

10

8

更新する(ステップ48)。

【0031】[実施例3]図9及び図10は、本発明の第3の実施例の処理フローを説明するための流れ図である。図9及び図10を参照して本発明の第3の実施例の動作について詳細に説明する。

【0032】本発明の第3の実施例における初期化手段2の処理は、図6に示した前記第2の実施例における初期化手段2と同じである。本発明の第3の実施例における予測手段3の処理は、図3に示した前記第1の実施例における予測手段3と同じである。

20

【0033】図9は、本発明の第3の実施例において、図3の重み合計sumの計算をするステップ34の処理フローを示す流れ図である。候補値aに投票する値の集合をV(a)とする。同じ列jで値が既に分かっている行i'の値O(i', j)がV(a)に含まれているような行i' と行iとの間の重みpow(2-γ, c(i, i')) · pow(γ, d(i, i'))の合計を、候補値aに投票する重みの合計sumとする(ステップ342-21)。但しpow(x, y)はxのy乗を表す。

20

【0034】ここで、計算に使うc(i, i'), d(i, i')は、図10に示されている処理に従って算出する。c(i, i')は行i'による行iへの投票が正しいと推測される証拠の量であり直接的な証拠の量C(i, i')と間接的な証拠の量との和である。

【0035】直接的な証拠は、ステップ342210で加算される。

30

【0036】間接的な証拠は、 $i'' \neq i$ かつ $i'' \neq i'$ であるようなすべての行i''について計算し加算する。先ず、行iと行i''との類似度 $g(i, i'') = C(i, i'') - D(i, i'')$ と、行i' と行i''との類似度 $g(i', i'') = C(i', i'') - D(i', i'')$ を計算する(ステップ342213)。

40

【0037】 $g(i, i'')$ と $g(i', i'')$ がともに非負であるとき(ステップ342214)のみ、間接的な証拠として $g(i, i'')$ と $g(i', i'')$ の絶対値の大きくない方をδ倍したものを $c(i, i'')$ に加える(ステップ342215)。但し、δは非負定数である。

【0038】 $d(i, i'')$ は行i'による行iへの投票が間違っていると推測される証拠の量であり直接的な証拠の量 $D(i, i'')$ と間接的な証拠の量との和である。

【0039】直接的な証拠は、ステップ342210で加算される。

【0040】間接的な証拠は、 $i'' \neq i$ かつ $i'' \neq i'$ であるようなすべての行i''について計算し加算する。先ず、行iと行i''との類似度 $g(i, i'') = C(i, i'') - D(i, i'')$ と行

50

9

i' と行 i'' との類似度 $g(i', i'') = C(i', i'') - D(i', i'')$ を計算する (ステップ342213)。

【0041】 $g(i, i')$ と $g(i, i'')$ の一方が正でもう一方が負であるとき (ステップ342216) のみ、間接的な証拠として $g(i, i')$ と $g(i, i'')$ の絶対値の大きさを α 倍したものを $d(i, i')$ に加える (ステップ342217)。

【0042】本発明の第3の実施例における学習手段4の処理フローは、図8に示した前記第2の実施例における学習手段4と同じである。

【0043】【実施例4】図11乃至図13は、本発明の第4の実施例の処理フローを説明するための流れ図である。図11及び図13を参照して本発明の第4の実施例の動作について詳細に説明する。

【0044】図11は、本発明の第4の実施例における初期化手段2の処理フローを示す流れ図である。前記第1の実施例における初期化手段2の流れ図 (図2) のステップ21、22、230、24、25を実行し、 $w(i, i')$ を「1」に $O(i, j)$ を「?」にセットする。さらに、任意の情報 j と情報 j' の間の重み $v(j, j')$ を「1」にセットする (ステップ280)。

【0045】本発明の第4の実施例における予測手段3の処理は、図3に示した前記第1の実施例における予測手段3の処理フローと同じである。

【0046】図12は、本発明の第4の実施例において、図3のsumの計算をするステップ34の処理を示す流れ図である。候補値 a に投票する値の集合を $V(a)$ とする。

【0047】前記第1の実施例における場合 (図4) と同様に、同じ列 j で値が既に分かっている行 i' の値 $O(i', j)$ が $V(a)$ に含まれているような行 i' と行 i の間の重み $w(i, i')$ の合計を、候補値 a に投票する重みの合計 sum とする (ステップ342011)。さらに、同じ行 i で値が既に分かっている列 j' の値 $O(i, j')$ が $V(a)$ に含まれているような列 j' と列 j の間の重み $v(j, j')$ の合計を、sum に加える (ステップ34202)。

【0048】図13は、本発明の第4の実施例における学習手段4の処理フローを示す流れ図である。前記第1の実施例における学習手段4の処理フローを示す流れ図である図5のステップ41、42、43、44、450、460、47を実行し、 $w(i, i')$ を更新する。

【0049】さらに、正解値 $M(i, j)$ に投票した列 j' ($O(i, j') \in V(M(i, j))$) と列 j の間の重み $v(j, j')$ は $(2-\gamma)$ 倍にし (ステップ4d0)、 $M(i, j)$ に投票しない他の値に投票

10

した列 j' ($O(i, j')$ は「?」でなくしかも $V(M(i, j))$ に含まれていない) と列 j の間の重み $v(j, j')$ は γ 倍にする (ステップ4e0)。但し、 γ は0以上1未満の定数である。重みの更新後、観測行列の (i, j) エントリの値 $O(i, j)$ を $M(i, j)$ に更新する (ステップ4f)。

【0050】【実施例5】図14乃至図16は、本発明の第5の実施例の処理フローを説明するための流れ図である。図14及び図16を参照して本発明の第5の実施例の動作について詳細に説明する。

【0051】図14は、本発明の第5の実施例における初期化手段2の処理フローを示す流れ図である。前記第2の実施例における初期化手段2の流れ図である図6のステップ21、22、231、24、25を実行し $C(i, i')$ と $D(i, i')$ を「0」にセットし、 $O(i, j)$ を「?」にセットする。

【0052】さらに、任意の情報 j と情報 j' の正しい投票回数 $E(j, j')$ と間違った投票回数 $F(j, j')$ を「0」にセットする (ステップ281)。

【0053】本発明の第5の実施例における予測手段3の処理フローは、図3に示した前記第1の実施例における予測手段3の処理と同じである。

【0054】図15は、本発明の第5の実施例において、図3のsumの計算をするステップ34の処理フローを示す流れ図である。

【0055】候補値 a に投票する値の集合を $V(a)$ とする。前記第2の実施例における場合 (図7) と同様に、同じ列 j で値が既に分かっている行 i' の値 $O(i', j)$ が $V(a)$ に含まれているような行 i' と行 i の間の重み $pow(2-\gamma, C(i, i'))$ $pow(\gamma, D(i, i'))$ の合計を、候補値 a に投票する重みの合計 sum とする (ステップ34211)。但し $pow(x, y)$ は x の y 乗を表す。

【0056】さらに、同じ行 i で値が既に分かっている列 j' の値 $O(i, j')$ が $V(a)$ に含まれているような列 j' と列 j の間の重み $pow(2-\gamma, E(j, j'))$ $pow(\gamma, F(j, j'))$ の合計を、sum に加える (ステップ34212)。

【0057】図16は、本発明の第5の実施例における学習手段4の処理フローを示す流れ図である。前記第2の実施例における学習手段4と同様に、図8のステップ41、42、43、44、451、461、47を実行し $C(i, i')$ と $D(i, i')$ を更新する。さらに、正解値 $M(i, j)$ に投票した列 j' ($O(i, j') \in V(M(i, j))$) と列 j の間の正しい投票回数を数えるカウンタ $E(j, j')$ に1を加え (ステップ4d1)、 $M(i, j)$ に投票しない他の値に投票した列 j' ($O(i, j')$ は「?」でなく、しかも $V(M(i, j))$ に含まれていない) と列 j の間の間違った投票回数を数えるカウンタ $F(j, j')$ に

11

「1」を加える(ステップ4e1)。重みの更新後、観測行列の(i, j)エントリの値O(i, j)をM(i, j)に更新する(ステップ48)。

【0058】[実施例6]図17及び図18は、本発明の第6の実施例の処理フローを説明するための流れ図である。図17及び図18を参照して本発明の第6の実施例の動作について詳細に説明する。

【0059】本発明の第6の実施例における初期化手段2の処理は、図14に示した前記第5の実施例における初期化手段2の処理フローと同じである。

【0060】本発明の第6の実施例における予測手段3の処理は、図3に示した前記第1の実施例における予測手段3の処理フローと同じである。前記第3の実施例における場合(図9)と同様に、同じ列jで値が既に分かれている行i'の値O(i', j)がV(a)に含まれているような行i' と行iとの間の重みpow(2-r, c(i, i'))pow(r, d(i, i'))の合計を、候補値aに投票する重みの合計sumとする(ステップ34221)。但しpow(x, y)はxのy乗を表す。ここで計算に使うc(i, i'), d(i, i')は、図10に示されている処理に従って算出する。

【0061】さらに同じ行iで値が既に分かれている列j'の値O(i, j')がV(a)に含まれているような列j' と列jとの間の重みpow(2-r, e(j, j'))pow(r, f(j, j'))の合計を、sumに加える(ステップ34222)。ここで計算に使うe(j, j'), f(j, j')は、図18に示されている処理に従って算出する。

【0062】e(j, j')は列j'による列jへの投票が正しいと推測される証拠の量であり、直接的な証拠の量E(j, j')と間接的な証拠の量との和である。

【0063】直接的な証拠は、ステップ342220で加算される。

【0064】間接的な証拠は、j' ≠ jかつj' ≠ j'であるようなすべての列j'について計算し加算する。

【0065】先ず、列jと列j'の類似度g(j, j') = E(j, j') - F(j, j')と、列j' と列jとの類似度g(j', j') = E(j', j') - F(j', j')を計算する(ステップ342223)。

【0066】g(j, j')とg(j', j')とともに非負であるとき(ステップ342224)のみ、間接的な証拠としてg(j, j')とg(j', j')の絶対値の大きくなかったものをe(j, j')に加える(ステップ342225)。但し、δは非負定数である。f(j, j')は列j'による列jへの投票が間違っていると推測される証拠の量であり直接的な証拠の量F(j, j')と間接的な証拠の

12

量との和である。

【0067】直接的な証拠は、ステップ342220で加算される。

【0068】間接的な証拠は、j' ≠ jかつj' ≠ j'であるようなすべての列j'について計算し加算する。先ず、列jと列j'の類似度g(j, j') = E(j, j') - F(j, j')と列j' と列j'の類似度g(j', j') = E(j', j') - F(j', j')を計算する(ステップ342223)。

【0069】g(j, j')とg(j', j')の一方が正でもう一方が負であるとき(ステップ342226)のみ、間接的な証拠としてg(j, j')とg(j', j')の絶対値の大きくなかったものをf(j, j')に加える(ステップ342227)。

【0070】なお、本発明の第6の実施例における学習手段4の処理フローは、図16に示した前記第5の実施例における学習手段4の処理フローと同じである。

20 【0071】

【発明の効果】以上説明したように、本発明によれば下記記載の効果を奏す。

【0072】本発明の第1の効果は、ユーザ間の重みが信頼度を考慮した値に設定される、ということである。

【0073】その理由は、本発明においては、共通に評価している情報の数が少ないユーザ間の重みは特別に大きくなったり小さくなったりすることがないからである。

【0074】本発明の第2の効果は、類似しているユーザ間とそうでないユーザ間の重みの差が、従来法よりも大きくなる、ということである。

【0075】その理由は、本発明においては、相関係数法よりも重みの取り得る値の幅が大きく、データの数の増加に伴って重みの差が大きくなっていくよう構成したためである。

【0076】本発明の第3の効果は、データの数が少ない時の予測性能を向上する、ということである。

【0077】その理由は、本発明においては、重みの更新がかけ算であるために、学習が速く、さらに同じユーザの他の情報に対する評価値も使ったり、同じ情報に対する評価値の分かっていないユーザ間の重みや、同じユーザに対する評価値の分かっていない情報間の重みを使う方法を行うことで、より多くの情報に基づいて予測しているからである。

【図面の簡単な説明】

【図1】本発明の実施の形態の構成を示すブロック図である。

【図2】本発明の第1の実施例における初期化手段の処理フローを示す流れ図である。

【図3】本発明の第1の実施例における予測手段を示す

流れ図である。

【図4】本発明の第1の実施例における予測手段のステップ34(図3)の詳細を示す流れ図である。

【図5】本発明の第1の実施例における学習手段の処理フローを示す流れ図である。

【図6】本発明の第2の実施例における初期化手段の処理フローを示す流れ図である。

【図7】本発明の第2の実施例における予測手段のステップ34(図3参照)の詳細を示す流れ図である。

【図8】本発明の第2の実施例における学習手段の処理フローを示す流れ図である。

【図9】本発明の第3の実施例における予測手段のステップ34(図3参照)の詳細を示す流れ図である。

【図10】本発明の第3の実施例における、図9のステップ34221の詳細を示す流れ図である。

【図11】本発明の第4の実施例における初期化手段の処理フローを示す流れ図である。

【図12】本発明の第4の実施例における予測手段のステップ34(図3参照)の詳細を示す流れ図である。

【図13】本発明の第4の実施例における学習手段の処理フローを示す流れ図である。

【図14】本発明の第5の実施例における初期化手段の処理フローを示す流れ図である。

【図15】本発明の第5の実施例における予測手段のステップ34(図3参照)の詳細を示す流れ図である。

【図16】本発明の第5の実施例における学習手段の処理フローを示す流れ図である。

【図17】本発明の第6の実施例における予測手段のステップ34(図3参照)の詳細を示す流れ図である。

【図18】本発明の第6の実施例における、図17のステップ34222の詳細を示す流れ図である。

【符号の説明】

1 入出力装置

2 初期化手段

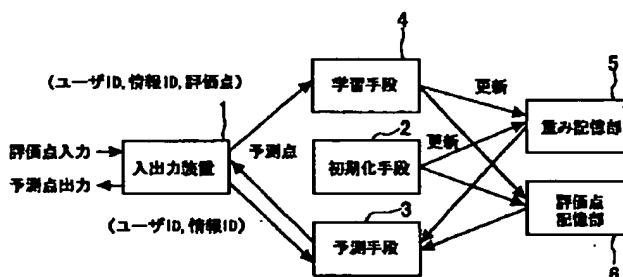
3 予測手段

4 学習手段

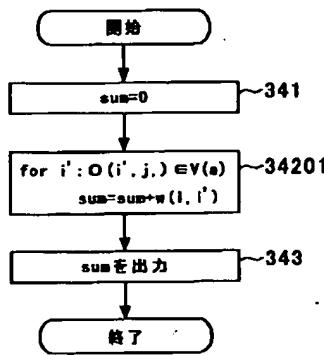
5 重み記憶部

6 評価点記憶部

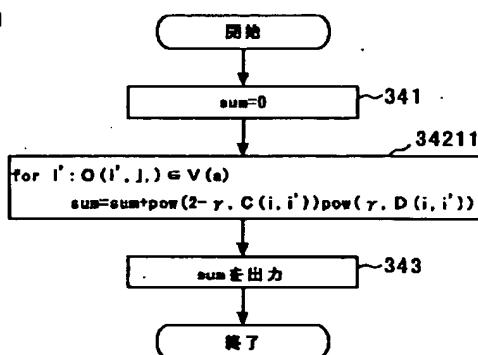
【図1】



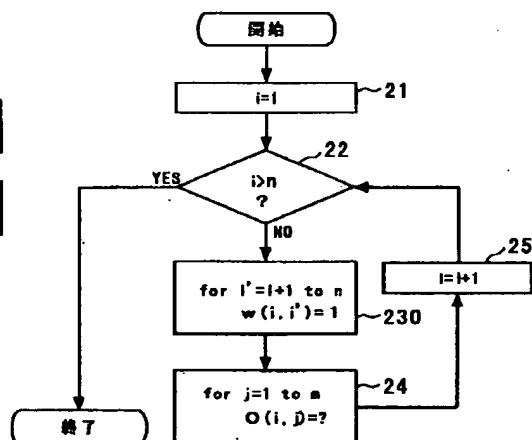
【図4】



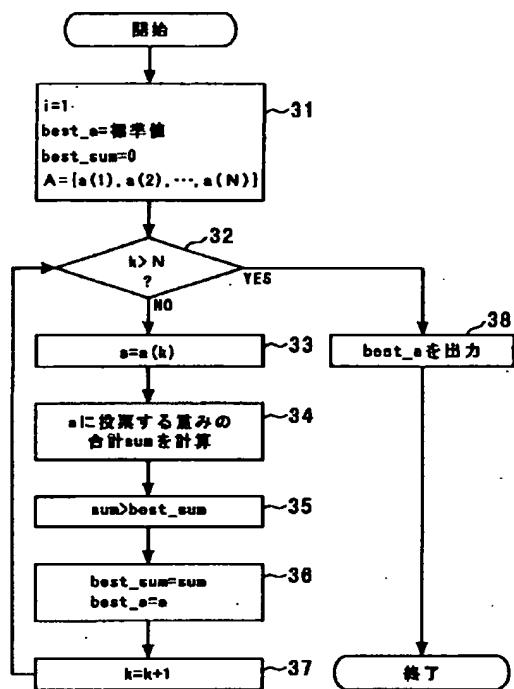
【図7】



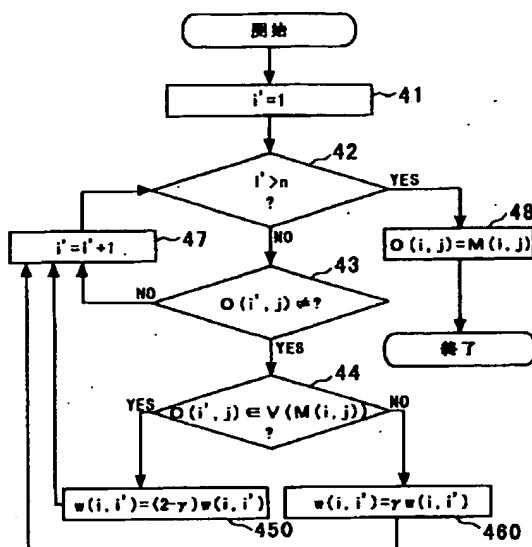
【図2】



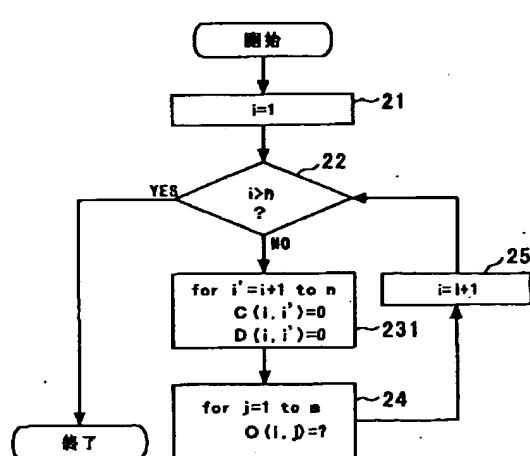
【図3】



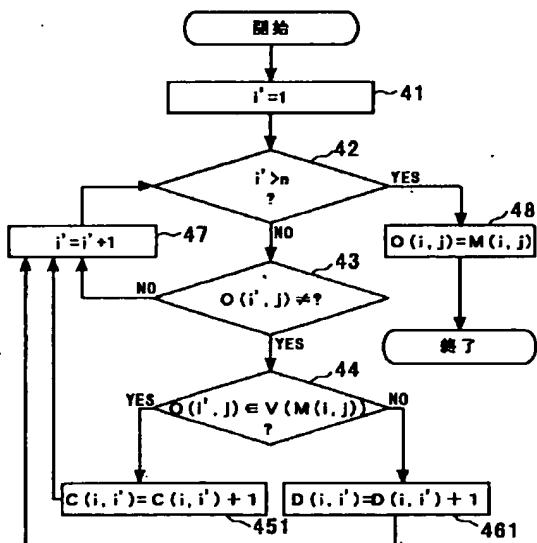
【図5】



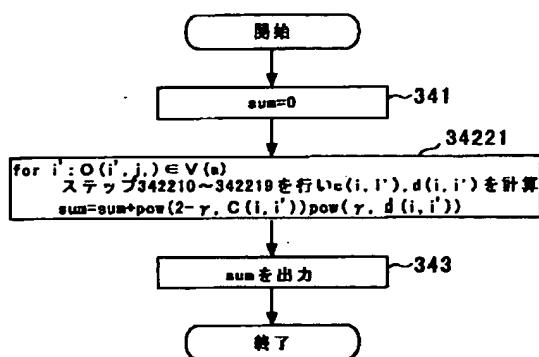
【図6】



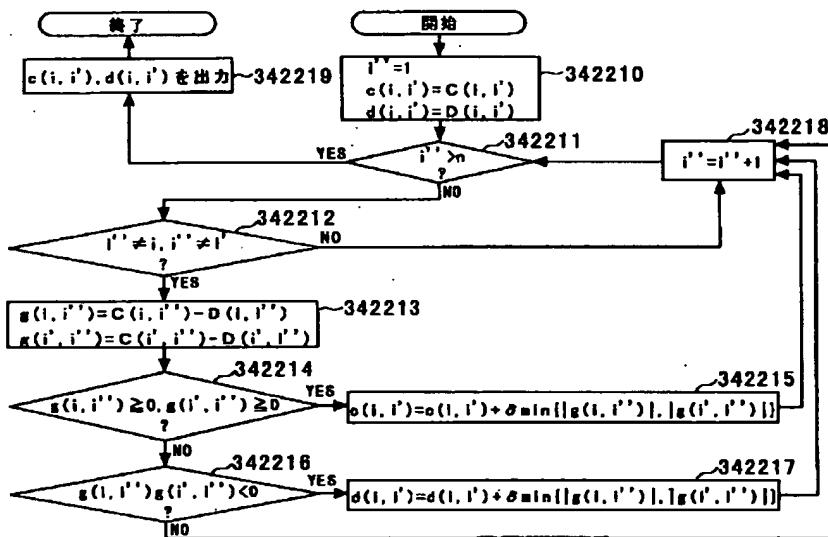
【図8】



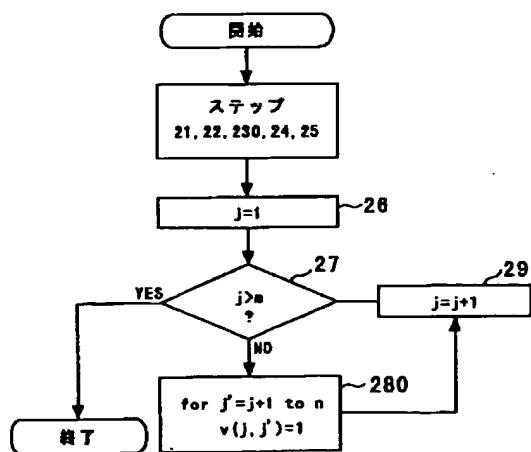
【図9】



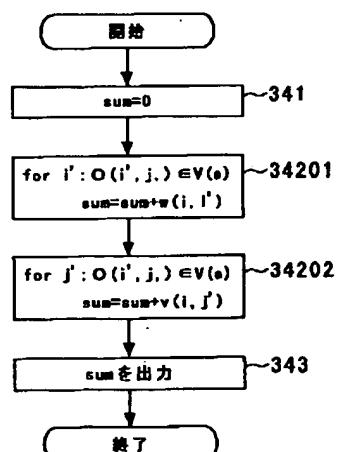
【図10】



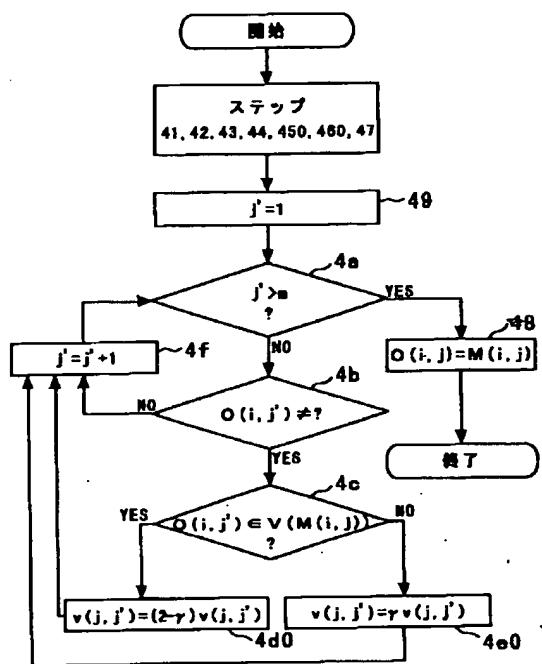
【図11】



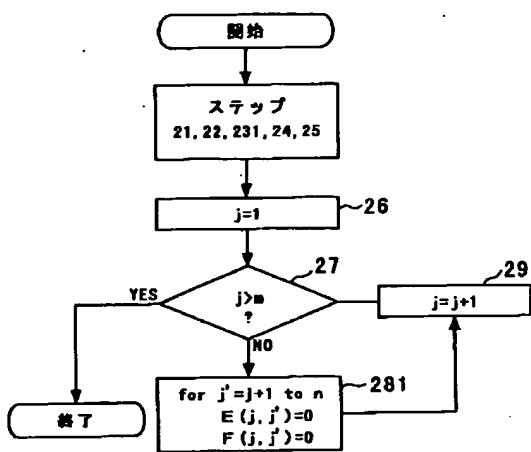
【図12】



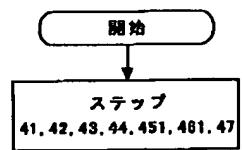
【図13】



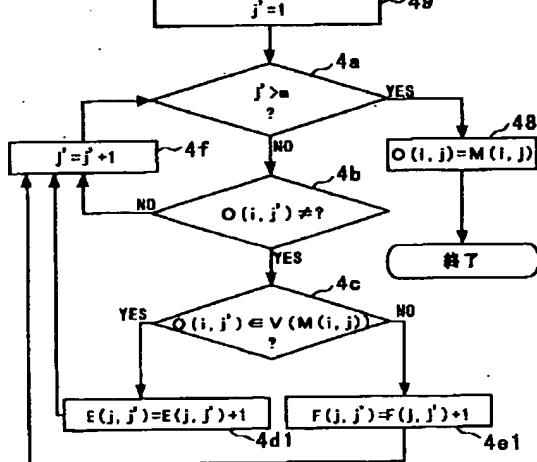
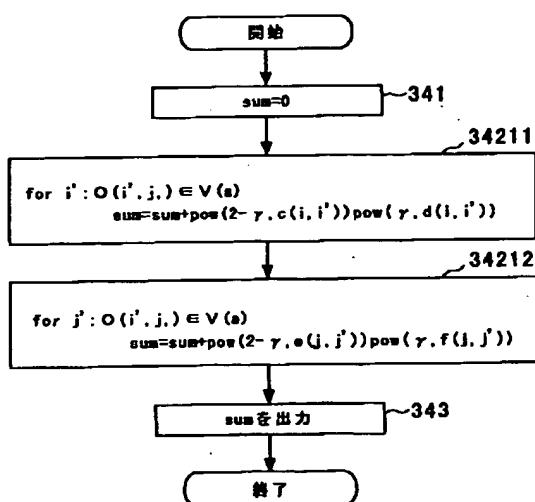
【図14】



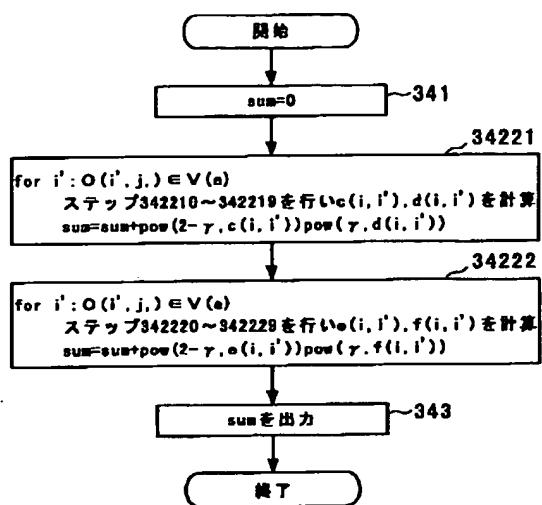
【図16】



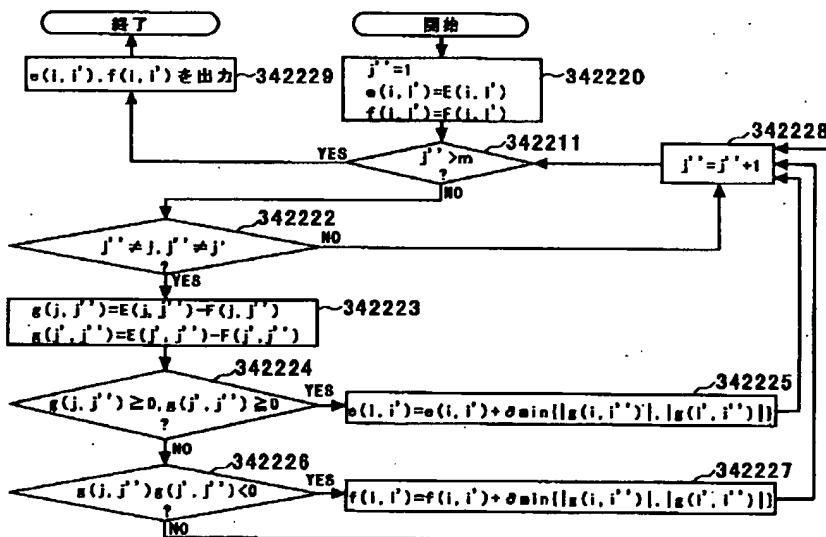
【図15】



【図17】



【図18】



(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平11-328275

(43)公開日 平成11年(1999)11月30日

(51)Int.Cl.
G 0 6 F 17/60
17/18

識別記号

F I
G 0 6 F 15/21
15/36

Z
Z

審査請求 有 請求項の数15 FD (全 12 頁)

(21)出願番号 特願平10-150593

(22)出願日 平成10年(1998)5月14日

(71)出願人 000004237

日本電気株式会社

東京都港区芝五丁目7番1号

(72)発明者 中村 勲祥

東京都港区芝五丁目7番1号 日本電気株式会社内

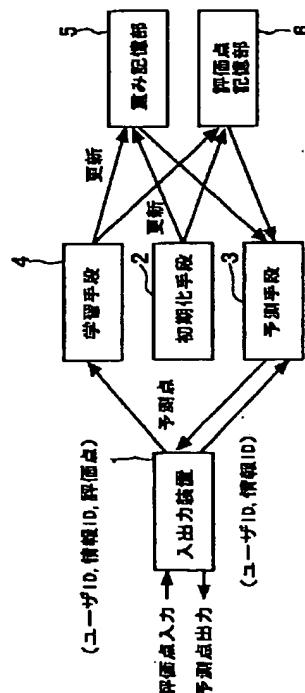
(74)代理人 弁理士 加藤 朝道

(54)【発明の名称】 情報フィルタリング方法及び方式

(57)【要約】

【課題】ユーザ間の重みを相関係数にしているため、信頼度を考慮し、類似しているユーザ間とそうでないユーザ間の重みの差がより大きくなるようにユーザ間の重み付けを可能とこれにより予測性能を向上する情報フィルタリング方法及び方式を提供。

【解決手段】各候補値に投票する値の集合を定義し、その集合に入る評価値のユーザとの間の重みの合計による多数決により予測値を決める予測手段3と、ユーザ間の重みは、正解に投票した数と不正解に投票した数からそのユーザ間の予測の正解度、誤り度を計算し、1に正解度の値だけ1より大きい数を掛け、誤り度の値だけ1より小さい数を掛けることにより計算し、ユーザ間の重みの学習は計算した値を記憶するか、正解数、誤り数を記憶することにより行う学習手段4を有する。



1

【特許請求の範囲】

【請求項1】ユーザーの情報に対する評価値の予測を、過去のユーザー、情報、評価値の3つ組のデータを使って学習したユーザ間の重みによる多数決により同じ情報を既に評価した他のユーザーの評価値から行う方法であって、各候補値に投票する値の集合を定義し、該集合に入る場合に、その候補値に投票する予測手段、を備えたことを特徴とする情報フィルタリング方式。

【請求項2】ユーザーの情報に対する評価値の予測を、過去のユーザー、情報、評価値の3つ組のデータを使って学習したユーザ間の重みと情報間の重みによる多数決により行う方法であって、

同じ情報を既に評価した他のユーザーの評価値のみでなく、同じユーザーの既に評価した他の情報の評価値も使い、各候補値に投票する値の集合を定義し、該集合に入る場合に、その候補値に投票する予測手段を備えたことを特徴とする情報フィルタリング方式。

【請求項3】正解値に投票する集合に入る場合には1よりも大きな所定の数を掛けることにより重みを増やし、入らない場合には1よりも小さな所定の数を掛けることにより重みを減らす学习手段を備えたことを特徴とする請求項1または2記載の情報フィルタリング方式。

【請求項4】重みを保持せず、予測の正解数と誤り数をそれぞれ計数する計数手段を備え、

前記計数手段のカウンタ値から重みを計算する予測手段と、

正解値に投票する集合に入る場合には正解数の計数手段の計数值を増やし、入らない場合には誤り数の計数手段の計数值を増やす学习手段と、

を備えたことを特徴とする請求項1または2記載の情報フィルタリング方式。

【請求項5】前記予測手段が、正解数と誤り数の計数手段からユーザー*i*の情報*j*と他のユーザー*i'*の情報*j'*の間の重みを計算する際に、そのユーザーの情報間の計数値のみでなく、他のすべてのユーザー*i''*の情報*j''*に対する前記ユーザー*i*の情報*j*との間の計数値、及び前記ユーザー*i'*の情報*j'*との間の計数値を使って、重みを算出することを特徴とする請求項4記載の情報フィルタリング方式。

【請求項6】ユーザーの情報に対する評価値の予測を、過去のユーザー、情報、評価値の3つ組のデータを使って学習したユーザ間の重みによる多数決により同じ情報を既に評価した他のユーザーの評価値から行う方法であって、各候補値に投票する値の集合を定義し、該集合に入る場合に、その候補値に投票することを特徴とする情報フィルタリング方式。

【請求項7】ユーザーの情報に対する評価値の予測を、過去のユーザーID、情報ID、評価値の3つ組のデータを使って学習したユーザ間の重みによる多数決により、同じ情報を、既に評価した他のユーザーの評価値から行う情

2

報フィルタリング方法であって、正解値を予測する際に、評価値の候補値毎に前記候補値に投票する重みの合計を計算し、重みの合計が最大となる候補値を出し、

正解値に投票する集合に入る場合には1よりも大きな所定の数を掛けることにより重みを増やし、入らない場合には1よりも小さな所定の数を掛けることにより重みを減らす、ことを特徴とする情報フィルタリング方法。

【請求項8】評価点の入力及び予測点の表示を行う入出力手段と、

重みを記憶する重み記憶部と、

評価点を記憶する評価点記憶部と、

前記重み記憶部及び評価点記憶部の初期化を行う初期化手段と、

与えられたユーザーIDと情報IDに対して予測点を出力する予測手段と、

入力評価データから学習して前記重み記憶部を更新し、前記評価点記憶部に入力データを加える学習手段と、を備え、

前記予測手段が、正解値を予測する際に、評価値の候補値毎に前記候補値に投票する重みの合計を計算し、重みの合計が最大となる候補値を出し、

前記学習手段が、正解値に投票する集合に入る場合には1よりも大きな所定の数を掛けることにより重みを増やし、入らない場合には1よりも小さな所定の数を掛けることにより重みを減らす、ことを特徴とする情報フィルタリングシステム。

【請求項9】前記予測手段において、任意のユーザー*i*とユーザー*i'*の間の重みをw(*i*, *i'*)とし、観測した

行列エントリ(ただし(*i*, *j*)成分がユーザー*i*による情報*j*の評価値となる)の値を記憶するための行列を観測行列O(*i*, *j*)とし、(*i*, *j*)エントリの正解値M(*i*, *j*)を予測する際に、候補値*a*に投票する値の集合をV(*a*)とし、同じ列*j*で値が既に分かっている行*i'*の値O(*i'*, *j*)が前記集合V(*a*)に含まれているような行*i'*と行*i*との間の重みw(*i*, *i'*)の合計を、候補値*a*に投票する重みの合計(s_{um})とし、

評価値の候補値*a*毎に*a*に投票する重みの合計(s_{um})を計算し、前記合計(s_{um})が最大となる候補値*a*を出力する、ことを特徴とする請求項8記載の情報フィルタリングシステム。

【請求項10】前記学習手段が、正解値M(*i*, *j*)に投票した行*i'*(O(*i'*, *j*) ∈ V(M(*i*, *j*)))と、行*i*との間の重みw(*i*, *i'*)は、(2- γ)倍し(ただし γ は0以上1未満の定数)、M(*i*, *j*)に投票しないで他の値に投票した行*i'*(O(*i'*, *j*)が未定でなく、V(M(*i*, *j*))に含まれていない)と行*i*との間の重みw(*i*, *i'*)は γ 倍

50 とし、重み更新後、観測行列の(*i*, *j*)エントリの値

3

$O(i, j)$ を、正解値 $M(i, j)$ に更新する、ことを特徴とする請求項8又は9記載の情報フィルタリングシステム。

【請求項11】前記予測手段において、任意のユーザ i とユーザ i' 間の正しい投票回数のカウンタ $C(i, i')$ と間違った投票回数のカウンタ $D(i, i')$ を用意し、

候補値 a に投票する値の集合を $V(a)$ とし、同じ列 j で値が既に分かっている行 i' の値 $O(i', j)$ が前記集合 $V(a)$ に含まれているような行 i' と行 i との間の重み $\text{pow}(2-r, C(i, i')) \cdot \text{pow}(r, D(i, i'))$ (ただし、 $\text{pow}(x, y)$ は x の y 乗の演算子) の合計を、候補値 a に投票する重みの合計とする、ことを特徴とする請求項8記載の情報フィルタリングシステム。

【請求項12】前記学習手段が、正解値 $M(i, j)$ に投票した行 i' ($O(i', j) \in V(M(i, j))$) と行 i との間の正しい投票回数を数えるカウンタ $C(i, i')$ に「1」を加え、前記正解値 $M(i, j)$ に投票しない他の値に投票した行 i' ($O(i', j)$ が未定でなくしかも集合 $V(M(i, j))$ に含まれていない) と行 i との間の間違った投票回数を数えるカウンタ $D(i, i')$ に「1」を加え、重みの更新後、観測行列の (i, j) エントリの値 $O(i, j)$ を $M(i, j)$ に更新する、ことを特徴とする請求項8又は11記載の情報フィルタリングシステム。

【請求項13】行 i' による行 i への投票が正しいと推測される証拠の量であり直接的な証拠の量 $C(i, i')$ と間接的な証拠の量との和である $c(i, i')$ と、行 i' による行 i への投票が間違っていると推測される証拠の量であり直接的な証拠の量 $D(i, i')$ と間接的な証拠の量との和である $d(i, i')$ を用い、重み $\text{pow}(2-r, c(i, i')) \cdot \text{pow}(r, d(i, i'))$ (ただし、 $\text{pow}(x, y)$ は x の y 乗の演算子) の合計を求める、ことを特徴とする請求項1記載の情報フィルタリングシステム。

【請求項14】前記間接的な証拠の量が、 $i'' \neq i$ かつ $i'' \neq i'$ であるようなすべての行 i'' について、先ず、行 i と行 i'' との類似度 $g(i, i'')$ = $C(i, i'') - D(i, i'')$ と、行 i' と行 i'' との類似度 $g(i', i'')$ = $C(i', i'') - D(i', i'')$ を計算し、 $g(i, i'')$ と $g(i', i'')$ がともに非負であるときのみ、間接的な証拠として $g(i, i'')$ と $g(i', i'')$ の絶対値の大きくなき方を所定倍数 δ (但し、 δ は非負定数) したものと加えることで求めることを特徴とする請求項11記載の情報フィルタリングシステム。

【請求項15】ユーザの情報に対する評価値の予測を、

4

過去のユーザ ID、情報 ID、評価値の3つ組のデータを使って学習したユーザ間の重みによる多数決により、同じ情報を、既に評価した他のユーザの評価値から行う情報フィルタリングシステムをコンピュータ上で実現するための処理であって、

(a) 正解値を予測する際に、評価値の候補値毎に前記候補値に投票する重みの合計を計算し、重みの合計が最大となる候補値を出力する処理、

(b) 正解値に投票する集合に入る場合には1よりも大きな所定の数を掛けることにより重みを増やし、入らない場合には1よりも小さな所定の数を掛けることにより重みを減らす処理、
の上記(a)、(b)の処理をコンピュータ上で実行する為のプログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は情報フィルタリング技術に関し、特にユーザの情報に対する評価値を他のユーザの情報に対する評価値から予測する協調フィルタリング技術に関する。

【0002】

【従来の技術】情報フィルタリングとは、データベース、パソコン通信、WWW (World Wide Web) 等情報システムが提供する大量の情報や多数のサービスの中から利用者の興味や嗜好に適うものを選択的に抽出するフィルタリング技術である。

【0003】従来より、ユーザの情報に対する評価値を他のユーザの情報に対する評価値から予測する協調フィルタリング (「Collaborative Filtering」または「Social Filtering」とも呼ばれる) としては、Proceedings of CSCW (1994)掲載の「GroupLens: An Open Architecture for Collaborative Filtering of News」やProceedings of CHI (1995)掲載の「Social Information Filtering: Algorithms and Automating "Word of Mouth"」に記載されているように、

ユーザ間の相関係数を重みとして既に評価済の他のユーザの評価値の重み付き平均により予測値を計算する方法が用いられている。

【0004】また、上記従来の方法における他のユーザの評価点が集中していると、集中している値の影響が大きすぎるという、問題点に対処すべく、例えば特開平9-265478号公報には、人口密度を考慮する情報フィルタリング方式が提案されている。

【0005】

【発明が解決しようとする課題】しかしながら、上記従来の情報フィルタリング方式は下記記載の問題点を有している。

【0006】第1の問題点は、例えば100個の情報を共通に評価していて、その100個に同じ評価をしているユーザ間の重みも、1個の情報を共通に評価している

だけでその1つが同じ評価であるユーザ間も、相関係数による重みは、同一となる、ということである。

【0007】その理由は、相関係数による方法は、信頼度を考慮していないからである。

【0008】第2の問題点は、特別評価が似ているユーザがいても、そのユーザとの間の重みが他と比べて特別高くはならない、ということである。

【0009】その理由は、相関係数の値の範囲は-1から1までであるからである。

【0010】第3の問題点は、データの数が少ない時の予測性能が悪い、ということである。 10

【0011】その理由は、上記第1、第2の問題点が解消されていないためである。

【0012】したがって本発明は、上記問題点に鑑みてなされたものであって、その目的は、信頼度を考慮し、類似しているユーザ間とそうでないユーザ間の重みの差がより大きくなるようにユーザ間の重み付けを可能とし、これにより予測性能を向上する情報フィルタリング方法及び方式を提供することにある。

【0013】

【課題を解決するための手段】前記目的を達成するため、本発明の情報フィルタリング方式は、ユーザの情報に対する評価値の予測を、過去のユーザ、情報、評価値の3つ組のデータを使って学習したユーザ間の重みによる多数決により、同じ情報を既に評価した他のユーザの評価値から行う方式であって、各候補値に投票する値の集合を定義し、該集合に入る場合にその候補値に投票する予測手段を備える。

【0014】また本発明の情報フィルタリング方式は、ユーザの情報に対する評価値の予測を、過去のユーザ、情報、評価値の3つ組のデータを使って学習したユーザ間の重みと情報間の重みによる多数決により行う方式であって、同じ情報を既に評価した他のユーザの評価値のみでなく同じユーザの既に評価した他の情報の評価値も使い、各候補値に投票する値の集合を定義し、該集合に入る場合にその候補値に投票する予測手段を備える。

【0015】さらに本発明においては、正解値に投票する集合に入る場合は1より大きい定数を掛けることにより重みを増やし、入らない場合には1より小さい定数を掛けることにより重みを減らす学習手段を備える。

【0016】

【発明の実施の形態】本発明の実施の形態について以下に説明する。本発明の情報フィルタリング方式は、その好みの実施の形態において、図1を参照すると、各候補値に投票する値の集合を定義し、その集合に入る評価値のユーザとの間の重みの合計による多数決により予測値を決める予測手段(3)を有する。

【0017】また、ユーザ間の重みは、正解に投票した数と不正解に投票した数からそのユーザ間の予測の正解度、誤り度を計算し、1に正解度の値だけ1より大きい

数を掛け、誤り度の値だけ1より小さい数を掛けることにより計算する。ユーザ間の重みの学習は計算した値を記憶するか、または、正解数、誤り数を記憶することにより行う学習手段(4)を有する。なお、重み記憶部及び評価点記憶部の初期化を行う初期化手段(2)、予測手段(3)、及び学習手段(4)は、コンピュータ上で実行されるプログラムによってその機能を実現することができる。本発明の実施の形態によれば、ユーザ間の重みを相関係数とし、信頼度を考慮し、類似しているユーザ間とそうでないユーザ間の重みの差が大となるようにユーザ間での重みの差が大となるようにユーザ間での重み付けを可能とし、少ないデータ数でも予測性能を向上している。

【0018】

【実施例】次に、上記した本発明の実施の形態についてさらに詳細に説明すべく、本発明の実施例について図面を参照して詳細に説明する。

【0019】図1は、本発明の一実施例の構成を示すブロック図である。図1を参照すると、本発明の一実施例は、評価点の入力及び予測点の表示を行う入力装置1と、重み記憶部5及び評価点記憶部6の初期化を行う初期化手段2と、与えられたユーザIDと情報IDに対して予測点を出力する予測手段3と、入力評価データから学習して、重み記憶部5を更新し、評価点記憶部6に入力データを加える学習手段4と、を含む。

【0020】【実施例1】図2乃至図5は、本発明の第1の実施例の処理フローを説明するための流れ図である。図1乃至図5を参照して、本発明の第1実施例の動作について詳細に説明する。

【0021】図2は、本発明の第1の実施例における初期化手段2の処理フローを示す流れ図である。任意のユーザ*i*とユーザ*i'* (*i'* は *i*+1~*n*, *i* は 1~*n*) の間の重み *w(i, i')* を「1」にセットし(ステップ230)、その時までに観測した行列エントリの値(*i, j*)成分はユーザ*i*による情報*j*の評価値)を記憶するために使う観測行列の任意のエントリ(*i, j*)の値 *O(i, j)* を「?」(未定)にセットする(ステップ24)。

【0022】図3は、本発明の第1の実施例における予測手段3の処理フローを示す流れ図である。*(i, j)*エントリの値 *M(i, j)* (正解値)を予測するとき、評価値の候補値 *a* 每に *a* に投票する重みの合計 *s um* を計算し(ステップ34)、*s um* が最大となる *a* を出力する(ステップ38)。

【0023】図4は、図3において、*s um* の計算をするステップ34の処理フローの詳細を示す流れ図である。候補値 *a* に投票する値の集合を *V(a)* とする。同じ列 *j* で値が既に分かっている行 *i'* の値 *O(i', j)* が *V(a)* に含まれているような行 *i'* と行 *i* の間の重み *w(i, i')* の合計を、候補値 *a* に投票する